

# Обучение машинного перевода без параллельных текстов\*

Артеменков<sup>1</sup> А. А., Бахтеев<sup>1</sup> О. Ю., Стрижов<sup>2</sup> В. В.

<sup>1</sup>Московский физико-технический институт

<sup>2</sup>Вычислительный центр им. А. А. Дородницына ФИЦ ИУ РАН

В данной работе исследуется задача машинного перевода между двумя языками. Предлагается подход, основанный на моделях автокодировщиков и не требующий наличия большого корпуса параллельных предложений. Каждому предложению из обоих языков ставится в соответствие вектор в общем скрытом пространстве. Оптимизация проводится таким образом, чтобы скрытые пространства автокодировщиков для разных языков совпадали. Для проверки качества модели проводится вычислительный эксперимент по переводу предложений между парой языков русский-украинский.

**Ключевые слова:** *нейронные сети, машинный перевод, автокодировщики.*

## Введение

Целью данной работы является решение задачи машинного перевода в отсутствии достаточного корпуса параллельных предложений. При наличии достаточно числа параллельных образцов (порядка нескольких миллионов [1]) хорошо себя показывают методы машинного перевода с использованием нейронных сетей ([2], [3]). Наилучшие результаты достигаются при использовании глубоких (свёрточных или рекуррентных) нейронных сетей, однако, в данном подходе критично наличие большой обучающей выборки. Частичное решение данной проблемы было найдено в пополнении числа предложений с помощью использования переводчиков более низкого качества. В [4] было показано, что данным способом могут быть улучшены результаты работы Moses ([5]). Более общим подходом является отказ от перевода в одну сторону и параллельное обучение переводчиков таким образом, чтобы один пополнял обучающую выборку другого.

Описанный выше метод был использован в [6] для перевода предложений с английского языка на французский. В данной работе подобная технология будет применяться для перевода с русского языка на украинский. Рассматриваются автокодировщики, реализованные в виде LSTM ([7], [8]), используемые для прямого и обратного перевода, и сеть-дискриминатор, обучаемая по представлению слова в векторном пространстве (например, [9]) определять язык. Автокодировщики подстраиваются таким образом, чтобы их латентные представления совпадали, или, что эквивалентно, чтобы дискриминатор не мог с достаточной уверенностью определить язык, соответствующий сгенерированному латентному вектору. Для того, чтобы избежать переобучения, добавляется шум, не дающий автокодировщикам переобучиться и начать восстанавливать предложения в точности. Шаг обучения состоит из двух стадий: обучение дискриминатора и обучение переводчика. На первой стадии выбирается случайное предложение из исходного языка, кодируется с добавлением шума (один из вариантов рассмотрен в [10]) и подаётся на вход дискриминатору. После подстройки его весов аналогичные действия повторяются со случайным предложением из конечного языка. На второй стадии выбирается случайное предложе-

ние из исходного языка и переводится текущей версией переводчика на конечный язык. Затем на него накладывается шум, оно кодируется, и считываются показания дискриминатора. После обновления весов предложение переводится обратно в исходный язык и вычисляется значение функции потерь. После шага оптимизации действия повторяются со случайным предложением из конечного языка. Качество полученного в результате переводчика оценивается с помощью метрики BLEU([11]).

## Литература

- [1] Bahdanau Dzmitry, Cho Kyunghyun, Bengio Yoshua. Neural machine translation by jointly learning to align and translate // arXiv preprint arXiv:1409.0473. 2014.
- [2] On the properties of neural machine translation: Encoder-decoder approaches / Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau [и др.] // arXiv preprint arXiv:1409.1259. 2014.
- [3] Luong Minh-Thang, Pham Hieu, Manning Christopher D. Effective approaches to attention-based neural machine translation // arXiv preprint arXiv:1508.04025. 2015.
- [4] Bertoldi Nicola, Federico Marcello. Domain adaptation for statistical machine translation with monolingual resources // Proceedings of the fourth workshop on statistical machine translation / Association for Computational Linguistics. 2009. С. 182–189.
- [5] Moses: Open source toolkit for statistical machine translation / Philipp Koehn, Hieu Hoang, Alexandra Birch [и др.] // Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions / Association for Computational Linguistics. 2007. С. 177–180.
- [6] Lample Guillaume, Denoyer Ludovic, Ranzato Marc'Aurelio. Unsupervised machine translation using monolingual corpora only // arXiv preprint arXiv:1711.00043. 2017.
- [7] Gers Felix A, Schmidhuber Jürgen, Cummins Fred. Learning to forget: Continual prediction with LSTM. 1999.
- [8] Graves Alex, Schmidhuber Jurgen. Framewise phoneme classification with bidirectional LSTM and other neural network architectures // Neural Networks. 2005. Т. 18, № 5-6. С. 602–610.
- [9] Goldberg Yoav, Levy Omer. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method // arXiv preprint arXiv:1402.3722. 2014.
- [10] Kim Yunsu, Ney Jiahui Geng Hermann. Improving Unsupervised Word-by-Word Translation with Language Model and Denoising Autoencoder. 2018.
- [11] BLEU: a method for automatic evaluation of machine translation / Kishore Papineni, Salim Roukos, Todd Ward [и др.] // Proceedings of the 40th annual meeting on association for computational linguistics / Association for Computational Linguistics. 2002. С. 311–318.